# A Study on the Analysis of Genetic Algorithms with Various Classification Techniques for Feature Selection

E.P.Ephzibah Research Scholar, Mother Teresa Women's University, Kodaikanal. India B. Sarojini Professor, Department of MCA, KLN College of Information Technology, MADURAI. J.Emerald Sheela Professor, Department of Mathematics, Madras Christian College, Chennai.

# ABSTRACT

In the world of curse of dimensionality feature selection plays a very important role in reducing the entire feature collection with the limited subset of features. Reducing the number of features pave way for various advantaged as well as simplifies the task. Feature selection means finding the suitable set of features which will contribute the most of it to the solution with minimal or null error rate. Selected features are to be tested with the help of classifiers, so that the subset of selected features can be proved to be optimal when compared to other features subsets individually as well as a group. Genetic algorithms are now days play a vital role among any other methodology in selecting the features based on the Theory of Evolution and on the "Survival of the fitness". It is a heuristic approach. To cooperate with the GA approach we have the classifiers which will go hand in hand to bring out the final set of features along with their prediction accuracy. In this paper I have analyzed four of the classifiers and compared them with their performance and the unit of accuracy.

### Keywords:

Feature Selection, Genetic Algorithm, SVM, KNN, Fuzzy Rough Set, Neural Networks, classification.

## **1. INTRODUCTION**

In pattern recognition and knowledge discovery the problem of feature selection has been widely investigated due to their importance. It allows the reduction of feature space, which is crucial in reducing the time that is taken for training and improving the accuracy of prediction. This can be obtained by removing the irrelevant, noisy and redundant features. In other words it is selecting the subset of features that can achieve better performance in prediction accuracy and computation time. The feature

selection algorithms include a few important components such as the initial population or the starting point in the feature space, search procedure, evaluation function or the fitness function and terminating condition. Initially all the features are taken into consideration. Later the subset of features can be found by evaluating all the possible solutions. These search procedures that are practical to implement are not guaranteed to find the optimal subset of features. Genetic algorithms is one of the search procedures which simulate natural evolution mechanisms of natural selection and natural genetics are used to in order to find solution to a problem. The basic operations involved in genetic algorithm are maintaining a population of solutions, selecting better solutions for recombination with each other and use their offspring to replace poorer solutions. GA is a combinatorial search technique based on random and probabilistic measures. Features are selected using a fitness function and then combined via cross-over and mutation operators to produce the next generation of subsets. To avoid local optima the population space is searched in parallel and not sequential. Results proved to be more efficient when the performance of GA is compared with other conventional methods.

# 2. DATA MINING AND CURSE OF DIMENSIONALITY

The dimensionality of the features in any application in the range of hundred is considered to be large. It can be hundreds of thousands in some applications. "Curse of dimensionality" is one of the major problems for feature selection. More efficient search strategies and evaluation criteria are needed for feature selection with large dimensionality.

### **3. FEATURE SELECTION**

Feature selection is a process that selects a subset of process should choose the optimal original features. The feature subset from a candidate set to describe the target conceptions of machine learning. The optimality of a feature subset is measured by an evaluation criterion. The number of features increases as the dimensionality expands. Problems that are related to dimensionality reduction and feature subset are found to be NP- hard. There are two models of feature selection that exist depending on whether the selection is coupled with a learning scheme or not. The first one is the filter model, which carries out the feature subset selection and the classification in two separate phases, uses a measure that is simple and fast to compute. This method is independent of the learning algorithm. This approach uses some metric function measuring a features' ability to distinguish the classes in data. The filter feature selection algorithm removes the irrelevant features without any learning algorithm. The drawback in this approach is that the feature subsets obtained by them may not be the best ones for a specific learning process. The second one is the wrapper method that performs feature subset selection and classification in the same process. It uses a learning algorithm as the metric function to find the accuracy of the classification.

#### **4. GENETIC ALGORITHM APPROACH:**

Genetic Algorithm is an adaptive heuristic method of global-optimization searching and simulates the behavior of the evolution process in nature. It maps the searching space into a genetic space. That is, every possible key is encoded into a vector called a chromosome. One element of the vector represents a gene. All of the chromosomes make up of a population and are estimated according to the fitness function. A fitness value will be used to measure the "fitness" of a chromosome. Initial populations in the genetic process are randomly created. GA then uses three operators to produce a next generation from the current generation: reproduction, crossover, and mutation. GA eliminates the chromosomes of low fitness and keeps the ones of high fitness. This whole process is repeated, and more chromosomes of high fitness move to the next generation, until a good chromosome (individual) is found. The main objective of genetic feature selection stage is to reduce the dimensionality of the problem before the supervised inductive learning process.

Among the many wrapper algorithms used, the Genetic Algorithm (GA), which solves optimization problems using the methods of evolution, specifically "survival of the fittest", has proved as a promising one. GA evaluates each individual's fitness as well the quality of the solution. The fitter individuals are more eligible to enter into the next generation as a population. After a required number of generations the final set of optimal population with fittest chromosomes will emerge giving the solution.

### **4.1 SELECTION:**

The reproduction process selects the chromosomes from the mating pool. The selection function chooses parents for the next generation based on their scaled values from the fitness scaling function. Roulette wheel selection is one common technique that implements the proportional selection strategy. It simulates a roulette wheel with the area of each segment proportional to its expectation. The algorithm then uses a random number to select one of the sections with a probability equal to its area. The selected sets of chromosomes are assumed to have fitter values for further operations.

### 4.2 CROSSOVER:

Crossover is a probabilistic process that exchanges information between two parent chromosomes for generating child chromosomes. Most commonly used crossover is the single point crossover with a fixed crossover probability of Pc is used. For chromosome of length l, a random integer called the crossover point c is generated. Single point chooses a random integer n between 1 and Number of variables, and selects the vector entries numbered less than or equal to n from the first parent, selects genes numbered greater than n from the second parent, and concatenates these entries to form the child. For example,

$$p1 = [a b c d e f g h]$$
  
 $p2 = [1 2 3 4 5 6 7 8]$ 

Random crossover point = 3

Child = [a b c 4 5 6 7 8] Crossover combines the two individuals to form a new individual for the next generation.

#### 4.3 MUTATION:

It is a process in which the chromosomes undergo mutation with a fixed probability Pm. In the chromosome a bit position (or gene value) is flipped to produce the new child. Mutation functions make small random changes in the individuals in the population, which provide genetic diversity and enable the Genetic Algorithm to search a broader space. For example,

$$p1 = [1 \ 0 \ 1 \ 1 \ 1 \ 1 \ 0 \ 1]$$

Random mutation point = 3

$$Child = [1 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 1]$$

Theses operators are found to improve the solution and also involve themselves with the operation that exists in nature. To obtain the best solution we need to measure each and every operator with a reasonable value. The selection or reproduction can be based on the "Roulette wheel selection", the crossover can have the value 0.8, mutation can have the value 0.002.

# 5. GENETIC ALGORITHM WITH OTHER CLASSIFICATION METHODS:

# 5.1 Genetic Algorithm and Support Vector Machine (SVM) Classifier:

The support vector machine and genetic algorithms is a prominent methodology that has received considerable attention in various application domains. Promising results have been obtained in medical diagnosis, text and image recognition etc... SVM is a supervised learning algorithm capable of solving linear and non-linear binary classification problems. Genetic algorithm is a meta- heuristic algorithm that solves mathematical optimization problem. They are based on Darwin's principle of "Survival of the fittest". Genetic algorithm has been used with SVM in several ways like feature selection, optimization of SVM parameters and kernel construction. GA based development of SVM is an iterative process starting with an initial population of randomly generated genotypes. GA provides excellent feature selection capabilities for any data set. It employs a population of competing solutions evolved over time by cross over mutation and selection to converge to an optimal solution. The specialty of genetic algorithm is that it produces a set of solutions instead of a single solution so as to avoid getting trapped in a local optimum. It is a promising method for feature selection over a high dimension space. The efficiency and accuracy of the SVM classifier are not only affected by the feature subset, but also the kernel function or the kernel parameters if the function has been specified (Weston et al., 2001). Moreover, feature subset and kernel parameters should be optimized simultaneously to get the optimal results (Huang & Wang, 2006).

In this section we will briefly describe the basic SVM concepts for typical two-class classification problems. These concepts can also be found in (Kecman, 2001; Schőlkopf and Smola, 2000; Cristianini and Shawe-Taylor, 2000). Given a training set of instance-label pairs ( $x_i$ ,  $y_i$ ), i=1, 2, ..., m where  $x_i \in \mathbb{R}^n$  and  $y_i \in$ 

$$\operatorname{Min}_{w,b,\xi} \frac{1}{2} w^{\mathrm{T}} w + C \sum_{i=1}^{m} \xi_i$$

subject to :  $y_i(\langle w \cdot x_i \rangle + b) + \xi_i - 1 \ge 0, \quad \xi_i \ge 0$ 

SVM provides a generic mechanism that fits the hyper plane surface to the training data using a kernel function. The user may select a kernel function (e.g. linear, polynomial, or sigmoid) for the SVM during the training process that selects  $\{+1, -1\}$ , for the linearly separable case, the data points will be correctly classified by

$$\langle w \cdot x_i \rangle + b \ge +1$$
 for  $y_i = +1$  (1)  
 $\langle w \cdot x_i \rangle + b \le -1$  for  $y_i = -1$  (2)

Eqs. (1) and (2) can be combined into one set of inequalities.

 $y_i(\langle w \cdot x_i \rangle + b) - 1 \ge 0 \quad \forall i = 1, ..., m$  <sup>(3)</sup>

The SVM finds an optimal separating hyper plane with the maximum margin by solving the following optimization problem:

 $\underset{i=1}{\text{Min}} \frac{1}{2} w^{T} w \text{ subject to } : y_{i} (\langle w \cdot x_{i} \rangle + b) - 1 \ge 0 \quad (4)$ 

In the case of linear generalized SVM, the goal is to construct a hyper plane that makes the smallest number of errors. To get a formal setting of this problem we introduce the non-negative slack variables  $\xi_i \ge 0$ , i=1,...,m. Such that

$$\langle w \cdot x_i \rangle + b \ge +1 - \xi_i \text{ for } y_i = +1$$
 (5)  
 $\langle w \cdot x_i \rangle + b \le -1 + \xi_i \text{ for } y_i = -1$  (6)

In terms of these slack variables, the problem of finding the hyperplane that provides the minimum number of training errors, i.e. to keep the constraint violation as small as possible, has the formal expression:

support vectors along the surface of this function. On an average it is found that the combination of genetic algorithm for feature selection and SVM for classification is found to be 90% efficient.

# 5.2 Genetic Algorithm and KNN Method for Feature Selection:

K -Nearest Neighbor algorithm is an instance based supervised learning that has been used in many applications of

data mining, statistical pattern recognition, image processing etc... KNN algorithm accomplishes very good performance in their experiments on different datasets. It is one of the most popular algorithms for pattern recognition. KNN algorithm is proved to accomplish good results in the experiments on different datasets. Combining genetic algorithm with KNN improves the classification performance. Genetic algorithm takes the k- neighbors and calculates the distance to classify the test samples instead of considering all the test samples.

In GA KNN method k the number of samples from the training set is generated and considered as initial population. The fitness value for the chromosomes is calculated so that the best fitness (highest fitness value) individuals are selected and they are stored. The most natural choice of the fitness function is some measure of the classification performance of the KNN rule. The value that obtained by calculating the distance between the best individuals is stored as global maximum. The process continues with a set of steps for a required no. of limited iterations.

The steps that are involved are:

- 1. Create a population.
- 2. Apply crossover and mutation operators.
- 3. Find the local minima.
- 4. Compare the value obtained for local minima with the existing global maximum.
- 5. The higher value is taken as global maximum.
- 6. The steps from 1-5 are repeated until the required number of iterations or the required global maximum.

Even though there are many limitations with KNN classifier, all of them can be combated by combining genetic algorithm approach. It is proved that the classification accuracy of GA KNN method is above 81%.





# **5.3 Genetic Algorithm and Fuzzy Sets for Feature Selection:**

A rough set is a formal approximation of a crisp set, in terms of a pair of sets which give lower and upper approximation of the original set. The lower and upper approximate sets are crisp sets. Fuzzification is a process in which the data is represented using a function called membership function. The quantitative value is transferred into fuzzy sets. The two membership values (yes or no) are produced for each attribute from the membership function.

Fuzzy rough sets are an extension for rough sets. They are applied in situations where classes are described as fuzzy sets on the feature space. Classification accuracy can also be obtained as a measure of how the data is classified in fuzzy recognition problems.

Fuzzy rough sets for feature selection play an important role in many applications where the data under consideration is not discrete. In a classical rough set theory it is possible to consider real valued data. Genetic algorithm can be employed to improve the fuzzy methods in order to achieve best classification precision. The inductive learning of a fuzzy rule-based classification system first of all determines a set of fuzzy rules from the set of instances and patterns. Each of these patterns is described by a set of features which are called as variables or characteristics. Genetic algorithm is used as a rule selection method that is used to get a co-operative set of rules. The GA approach defines a limited set of features based on the concept "survival of the fittest" and provides the features to the fuzzy system for further classification. The combination of feature selection process with the fuzzy rule generation and the genetic algorithm allows generating a set of variables which consider some characteristics of the fuzzy rule and the learning method.

# **5.4 Genetic Algorithm and Neural Networks for Feature Selection:**

Artificial Neural Networks can be used successfully in many areas of research in machine learning. ANN faces a problem when it has to take the inputs i.e., in the selection of best inputs which allows the creation of compact, high accurate networks which requires comparatively little preprocessing. The genetic algorithm is used to select the most significant input feature subset from a large dimensionality problem space.

Genetic algorithm with the help of fitness function and the operators like selection, crossover and mutation could select optimal feature subset.

Research on the combined work of genetic algorithm with artificial neural network is better that a stand alone artificial neural network. It is proved that genetic algorithm combined with the neural networks for feature selection and classification can give at the most of 100% accuracy.

### 6. RESULT AND CONCLUSION

The graphical representation given below depicts the performance of all the four combinations with respect to the accuracy of the model. It is clearly shown that the accuracy in the case of genetic algorithms with neural networks is the highest.



Graph: 1. Performance Relationship curve of the four combinations of genetic algorithm and the classifiers.

Genetic algorithm is an evolutionary algorithm that is capable of selecting the features for classification. This paper focuses on the performance of the classifiers for the same data and the analysis shows that the combination of genetic algorithms with the neural network classifier proves to be the best when compared with other classifiers.

#### 7. **REFERENCES**

1. Bengio Y., 2000. Gradient-based Optimization of Hyperparameters.Neural Computation, 12(8), pp. 1889-1900.

2. Bi J., Bennett K.P., Embrechts M., Breneman C.M., Song M., 2003. Dimensionality reduction via sparse support vector machines. *J. Mach. Learn. Res.* 3, pp. 1229-1243.

**3.** CHEN Pengwai, WNAG Jungying, LEE Hahnming " Model selection of SVMs with GA Approach" Proc 2004 IEEE Int joint conf on Neural.

4. Cortes C., Vapnik V.N., 1995. Support vector networks. Mach. Learn. 20(3), pp. 273-297.

5. Dasarathy.B.V, Nearest Neighbor (NN) Norms: NN Pattern classification Techniques, IEEE Computer society press, Los Alamitos California (1990)

6. Guyon I., Weston J., Barnhill S., Bapnik V., 2002. Gene selection for cancer Classification using support vector machines. *Machine Learning*, 46(1-3), pp. 389-422.

7. Jacek Jarmulak and Susan craw, "Genetic algorithms for Feature Selection and Weighting" Proc of the IJCAI '99 Workshop on Automating the construction of case based reasoners.

8. Jensen.R, Q. Shen, Tolerance-based and fuzzy-rough feature selection, in: Proceedings of the 16th International Conference on Fuzzy Systems (FUZZ- IEEE'07), 2007, pp. 877–882.

9. K. Kira, L.A. Rendell, The feature selection problem: traditional methods and a new algorithm, in: Proceedings of Ninth National Conference on Artificial Intelligence, 1992, pp. 129–134.

10. Klir, G.J. and B.yuan (1995) fuzzy sets and fuzzy logic theory and applications, Prentice Hall, Englewood cliffs, NJ, USA.

11. Metin. N., Heang-Ping, C., and Berhan, S., Optimal Neural Network Architecture Selection: Improvement in Computerized Detection of Micro calcifications, Acad Radiol 9, 2002, pp.420-429.

12. O. Maimon, A. Kandel, and M. Last, Information-Theoretic Fuzzy Approach to Knowledge Discovery in Databases. In Advances in Soft Computing - Engineering Design and Manufacturing, R. Roy, T. Furuhashi and P.K. Chawdhry, Eds. Springer- Verlag, London, 1999.

13. Raymer M.L., PunchW.F., 2000. Goodman E.D., Kuhn L.A., Jain A.K. Dimensionality reduction using genetic algorithms. *IEEE Trans. Evol. Comput.*, 4(2), pp. 164-171

14. Ruck D.W. S.K.Rogers and M.Kabrisky(1990) Feature Selection using a multilayer perceptron", Journal of Neural Network Computing. 40-48

15. Sikora R., Piramuthu S., 2007. Framework for efficient feature selection in genetic algorithm based data mining. European *Journal of Operational Research*, 180(2), pp. 723-737.

16. Swiniarsk.R.W.i, A. Skowron, Rough set methods in feature selection and recognition, Pattern Recognition Lett. 24 (6) (2003) 833–849.

17. Verma, B. K., A neural network based technique to locate and classify micro calcifications in digital mammograms, IEEE World Congress on Computational Intelligence, WCCI'98, Anchorage, USA, 1998, pp. 2163-2168.

18. Weston J., Mukherjee S., Chapelle O., Ponntil M., Poggio T., Vapnik V., 2001. *Feature selection for SVMs. Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA.

19. Zhong.N, J. Dong, S. Ohsuga, Using rough sets with heuristics for feature selection, J. Intell. Inf. Syst. 16 (3) (2001) 199–214.

20. Zilberstein.S Using Anytime Algorithms in Intelligent Systems, AI Magazine, vol. 17, no. 3, pp. 73-83, 1996